Econometric Theory, **17**, 2001, 1082–1112. Printed in the United States of America.

LIKELIHOOD-BASED INFERENCE IN TRENDING TIME SERIES WITH A ROOT NEAR UNITY

ZHIJIE XIAO
University of Illinois at Urbana-Champaign

This paper studies likelihood-based estimation and tests for autoregressive time series models with deterministic trends and general disturbance distributions. In particular, a joint estimation of the trend coefficients and the autoregressive parameter is considered. Asymptotic analysis on the *M*-estimators is provided. It is shown that the limiting distributions of these estimators involve nonlinear equation systems of Brownian motions even for the simple case of least squares regression. Unit root tests based on *M*-estimation are also considered, and extensions of the Neyman–Pearson test are studied. The finite sample performance of these estimators and testing procedures is examined by Monte Carlo experiments.

1. INTRODUCTION

In the past decade, econometricians have focused a great deal of attention on the development of estimation and hypothesis testing procedures in autoregressive time series models where the largest root is near unity. Most of these procedures are based on least square methods in linear regression models and have likelihood interpretations when the data are Gaussian. In the absence of Gaussianity, asymptotic results of these procedures generally still hold, but these methods are less efficient than methods that exploit the distributional information. Monte Carlo evidence indicates that the least squares estimator can be very sensitive to certain type outliers and that inference procedures based on the least square estimation may have poor performance (see, e.g., Lucas, 1994). In empirical analysis, many applications in nonstationary time series involve financial data such as exchange rates whose distributions are heavy-tailed. It is therefore important to consider estimation procedures that are robust to departures from Gaussianity and can be applied to nonstationary time series. The present paper addresses some of these issues.

This paper was presented at the Cowles Conference "New Developments in Time Series Econometrics" at Yale University in October 1999. Special thanks go to Peter Phillips for his encouragement and guidance. My thanks also go to Katsuto Tanaka, Anil Bera, Ted Juhl, Roger Koenker, Lung-Fei Lee, Francesc Marmol, Joon Park, and the referee for their helpful comments. The paper was partially supported by the Hong Kong Research Grants Council under grant No. CUHK 4078/98. Address correspondence to: Zhijie Xiao, Department of Economics, University of Illinois at Urbana-Champaign, 1206 South Sixth Street, Champaign, IL 61820, USA; e-mail: zxiao@uiuc.edu.

 © 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

There has been some study on nonstationary time series regression with nonnormal innovations, including Cox and Llatas (1991), Knight (1991), Phillips (1995), Lucas (1995), Rothenberg and Stock (1997), and Juhl (1999), among others. In particular, Phillips (1995) studies robust cointegrating regressions and develops fully modified least absolute deviations (LAD) and *M*-estimators for cointegrating regressions. Lucas (1995) considers unit root tests based on *M*-estimators. Cox and Llatas (1991) and Rothenberg and Stock (1997) study robust estimation and inference for nearly integrated autoregressive models without deterministic trends. In the multivariate case, Juhl (1999) proposes a test for cointegration using *M*-estimators based on singular value decomposition.

Many macroeconomic time series, such as real gross national product (GNP), consumption, money, and prices, display a tendency toward growth over time. Consequently, most empirical analyses in nonstationary time series literature consider unit root or near unit root processes with deterministic trends. One traditional way (see, e.g., Park and Phillips, 1988, 1989) of modeling trending time series is to consider regressions of the following form:

$$y_t = \gamma' x_t + \alpha y_{t-1} + u_t, \tag{1.1}$$

where x_t is a deterministic trend of known form. The high persistency in macroeconomic time series indicates that the autoregression coefficient α is close to 1, and there have been many empirical applications that test the null hypothesis of a unit root ($\alpha = 1$) against the alternative of stationarity ($\alpha < 1$). However, as argued in Schmidt and Phillips (1992), the parameterization in (1.1) is not convenient in interpreting the deterministic component. For instance, considering the leading case that $x_t = (1, t)'$, we have

$$y_t = \gamma_0 + \gamma_1 t + \alpha y_{t-1} + u_t. \tag{1.2}$$

Such an equation has the property that the meanings of the parameters γ_0 and γ_1 differ under the null and the alternative. Under the null of a unit root, the parameters γ_0 and γ_1 represent trend and quadratic trend, respectively. However, under the alternative, γ_0 and γ_1 determine level and trend. This problem also surfaces in the unit root tests, and an "extra" deterministic trend component has to be introduced to remove the nuisance parameters. The introduction of surplus trend variables results in some inefficiency in the regression and reduces the power of the corresponding unit root test from its already low level. To avoid the problem caused by the confusion over the meanings of parameters, researchers have considered as an alternative to (1.1) the following data generating process:

$$y_t = \gamma' x_t + y_t^s, \qquad t = 1, \ldots, n, \tag{1.3}$$

$$y_t^s = \alpha y_{t-1}^s + u_t, \qquad t = 1, \ldots, n. \tag{1.4}$$

This parameterization allows for the same trend component under both the null and the alternative hypothesis and is now widely used in time series analysis.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Combining (1.3) and (1.4) gives the nonlinear regression model (Phillips and Xiao, 1998)

$$y_t = \gamma'\Delta x_t + (1-\alpha)\gamma' x_{t-1} + \alpha y_{t-1} + u_t. \quad \textbf{(1.5)}$$

Compared with (1.1), regression (1.5) incorporates both the null and the alternative models in a nonlinear equation.

This paper considers likelihood-based estimation and hypothesis tests for autoregressive time series model (1.3) with deterministic trends and general disturbance distributions. In particular, we consider a joint estimation of the trend coefficients and the autoregressive coefficient based on the nonlinear regression (1.5). Asymptotic analysis of the *M*-estimators, including the maximum likelihood estimators, for both the trend and autoregression coefficients is provided. It is shown that the asymptotic distributions of these estimators are complicated and involve nonlinear equation systems of Brownian motions even for the simple case of least squares regression. We also consider unit root tests against local alternatives based on these estimators. Local power analysis is conducted to show that these tests have nontrivial power against *n*-local alternatives. In addition, as a natural extension of the Neyman–Pearson test, the likelihood ratio test for a unit root against a point alternative is studied, and asymptotic power functions and power envelopes are derived. Parallel to the existing study on the Gaussian case, unit root tests based on *M*-estimation coupled with quasi-differencing are analyzed. A Monte Carlo experiment is conducted and shows that these estimators display rather good finite sample properties when the data density has a heavy tail.

The paper is organized as follows. Section 2 describes estimation. The joint estimation of the trend coefficients and the local parameter is discussed, and asymptotic analysis of the *M*-estimators is given. Section 3 describes unit root tests based on these *M*-estimators, including likelihood ratio tests for a unit root against a general local alternative, the case with a point alternative, and the quasi-differencing *M*-detrended unit root tests. Monte Carlo results on the finite sample performance of the nonlinear estimators and associated tests are reported in Section 4, and Section 5 concludes. All of the proofs appear in the Appendix.

A word on notation: the symbol $\Rightarrow$ signifies weak convergence, $\equiv$ signifies equality in distribution, and $:=$ signifies definitional equality. L denotes lag operator. The expression $\Delta = 1 - L$ is the difference operator, and Δ_c signifies quasi-difference, which is defined by $\Delta_c = 1 - (1 + c/n)L$. The term $I(k)$ denotes integration of order k. All limits are taken as $T \to \infty$, unless otherwise specified.

2. ESTIMATION

2.1. Joint Estimation of the Trend and the AR Coefficient

Consider the autoregression model introduced in Section 1 in which the observed time series y_t can be written as the sum of a deterministic trend d_t and a stochastic component y_t^s:

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

$$y_t = d_t + y_t^s, \qquad t = 1, \ldots, n, \tag{2.1}$$

$$y_t^s = \alpha y_{t-1}^s + u_t, \qquad t = 1, \ldots, n. \tag{2.2}$$

The deterministic trend d_t depends on unknown parameters and is specified as

$$d_t = \gamma' x_t, \tag{2.3}$$

where γ is a vector of the trend coefficient and x_t is a deterministic trend of known form. The leading case of the deterministic component is the linear time trend where $x_t = (1, t)$. In general, the trend function x_t may be more complex than a simple time polynomial. For example, time polynomials with sinusoidal factors and piecewise time polynomials may be used. The latter corresponds to a class of models with structural breaks in the deterministic trend. The term y_t^s is the stochastic component of y_t and can be represented by an autoregressive process. The expression $\{u_t\}$ is the unobserved innovation process, which is assumed to be stationary with mean zero. We also assume for convenience that the initial observation y_1^s is a constant (more generally, without affecting the asymptotic results, we can assume that it is a random variable of finite variance).

Our purpose is to study likelihood based inference in this model when the autoregressive parameter α is closed to one. To obtain large sample approximations, we employ the local-to-unity asymptotic theory investigated by Phillips (1987, 1988), Chan and Wei (1987), and others. Thus we consider the parameter space in a shrinking neighborhood of unity and reparameterize α so that

$$\alpha = 1 + \frac{c}{n}. \tag{2.4}$$

Combining (2.1)–(2.4), we have the following nonlinear regression:

$$\Delta y_t = \gamma' \Delta x_t - c\gamma' \left(\frac{x_{t-1}}{n} \right) + c \left(\frac{y_{t-1}}{n} \right) + u_t. \tag{2.5}$$

To simplify the exposition, we consider the simple case that u_t are unobserved independent and identically distributed (i.i.d.) errors with mean zero and unit variance (for discussions, see Remark 6, which follows). If we denote the log density of u as $f(u)$, then the conditional log density of y_t is given as follows:

$$f(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n)),$$

and the joint log density of the random sample is

$$\sum_{t=2}^{n} f(\Delta y_t - c(y_{t-1}/n) - \gamma' \Delta x_t + c\gamma'(x_{t-1}/n)).$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Writing it as a function of the parameters, the preceding expression delivers the log likelihood function, and we denote it as $L(c,\gamma)$:

$$L(c,\gamma) = \sum_{t=2}^{n} f(\Delta y_t - c(y_{t-1}/n) - \gamma'\Delta x_t + c\gamma'(x_{t-1}/n)).$$

The maximum likelihood estimators of c and γ can then be found by maximizing $L(c,\gamma)$ with respect to c and γ. More generally, if we consider some criterion function φ, the so-called M-estimators of c and γ are obtained from a similar optimization problem with f replaced by φ.

Lucas (1995) considers unit root tests based on the M-estimator of model (1.1). Cox and Llatas (1991) study the asymptotic behavior of the M-estimator of c for the case without a deterministic trend, and Rothenberg and Stock (1997) consider model (1.3), but the asymptotic analysis was only conducted for the simple case without deterministic trends. However, in the presence of an unknown deterministic component, the system that determines the maximum likelihood estimator becomes more complicated and is generally nonlinear.

2.2. Asymptotic Analysis of the *M*-Estimators

We are interested in the asymptotic behavior of the M-estimators of c and γ in regression (2.5), defined as the solution of the following extreme problem:

$$\begin{bmatrix} \tilde{c} \\ \tilde{\gamma} \end{bmatrix} = \arg\max \left[\sum_{t=2}^{n} \varphi(\Delta y_t - c(y_{t-1}/n) - \gamma'\Delta x_t + c\gamma'(x_{t-1}/n)) \right] \tag{2.6}$$

for some criterion function φ. Taking $\varphi(u) = -u^2$, (2.6) gives the ordinary least squares (OLS) estimator of c and γ. The maximum likelihood estimator corresponds to the case when φ is the true log density function. Although we pay particular attention to the maximum likelihood estimator, our analysis in this section will be given in a general way so that the M-estimator is covered, treating the maximum likelihood estimator as a special case of particular interest (notice that some simplifications happen in the asymptotic results when φ is the true log density).

We want to examine the asymptotic distribution of the estimator $(\tilde{c},\tilde{\gamma})$. Under regularity conditions, the estimator $(\tilde{c},\tilde{\gamma})$ can also be defined as a solution to the following equation system, which is the first-order condition of the extremum problem (2.6):

$$(NL)\begin{cases} \sum_{t=2}^{n} \varphi'(\Delta y_t - c(y_{t-1}/n) - \gamma'\Delta x_t + c\gamma'(x_{t-1}/n))\Delta_c x_t = 0, \\ \sum_{t=2}^{n} \varphi'(\Delta y_t - c(y_{t-1}/n) - \gamma'\Delta x_t + c\gamma'(x_{t-1}/n))y_{t-1}^{s} = 0. \end{cases}$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

The expression (NL) is a nonlinear equation system and generally has no analytic solution even for the simple case when φ is the log density of a normal distribution. Indeed, as will become clear later, in a n-shrinking neighborhood of unity, the limiting distributions of the M-estimators are jointly determined by a nonlinear equation system of Brownian motions and the limiting trending functions.

To study the asymptotic distributions of the M-estimators, it is convenient for us to make the following assumptions on u_t and the criterion function φ.

Assumption A. u_t are i.i.d. with mean zero and variance one. The term $\varphi(\cdot)$ possesses derivatives φ' and φ''. Here $[u, \varphi'(u)]$ has kth moments for some $k > 2$, $E[\varphi'(u_t)] = 0$, and φ'' is Lipschitz continuous.

Assumption A is a standard condition in asymptotic analysis of maximum likelihood estimators or M-estimators. The assumption that u_t has unit variance is just for simplicity of exposition and brings no loss of generality. (In the general case with variance σ^2, a similar result can be obtained. The analysis given in this paper remains valid after a simple restandardization by σ^2; see Remark 6 for a discussion on the more general case.) The moment conditions on u and $\varphi'(u)$ are needed to establish the weak convergence results. We may also replace the moment condition on $\varphi'(u)$ by boundedness conditions of the derivatives of φ, because the latter and the moment condition on u imply the corresponding condition on φ'.

Denote $[\cdot]$ as the greatest lesser integer function; then under Assumption A, as n goes to ∞, $n^{-1/2}\sum_1^{[nr]} u_t$ converges weakly to a standard Brownian motion $W_1(r)$, and thus $n^{-1/2}y^s_{[nr]}$ converges weakly to the corresponding Ornstein–Uhlenbeck process $J_c(r) = \int_0^r e^{c(r-s)}dW_1(s)$. The limiting distributions of $\tilde{c}$ and $\tilde{\gamma}$ will also be dependent on the weak limit of the partial sums of $\varphi'(u_t)$. We denote $\omega^2 = \mathrm{var}[\varphi'(u_t)]$, $\delta = -E[\varphi''(u_t)]$, and $\rho = -E[u_t\varphi'(u_t)]$; then $n^{-1/2}\sum_1^{[nr]}\varphi'(u_t) \Rightarrow B_\varphi(r) = \omega W_\varphi(r)$, where W_φ is a standard Brownian motion.

Notice that $W_1(r)$ and $W_\varphi(r)$ are correlated Brownian motions. To deal with the correlation between W_1 and W_φ explicitly, following the previous literature, we construct the random variable $v_t = \varphi'(u_t) + \rho u_t$. Then, by construction, v_t are i.i.d. with variance $(\omega^2 - \rho^2)$ and are uncorrelated with u_t. The partial sum process $n^{-1/2}\sum_1^{[nr]} v_t$ converges weakly to $\sqrt{\omega^2 - \rho^2}W_2(r)$, where $W_2(r)$ is a standard Brownian motion independent of $W_1(r)$. The Brownian motion $B_\varphi(r)$ then has the following decomposition: $B_\varphi(r) = \omega W_\varphi(r) = \sqrt{\omega^2 - \rho^2}W_2(r) - \rho W_1(r)$.

For asymptotic analysis of the deterministic trend, we assume that there are standardizing matrices D_n and $F_n = n^{-1}D_n$ such that $D_n^{-1}x_{[nr]} \to X(r)$ and $F_n^{-1}\Delta x_{[nr]} \to g(r)$, as $n \to \infty$, uniformly in $r \in [0,1]$, where $X(r)$ and $g(r)$ are limiting trend functions. In the case of a linear trend, $D_n = \mathrm{diag}[1, n]$ and $X(r) = (1, r)'$. If x_t is a general pth order polynomial trend, $D_n = \mathrm{diag}[1, n, \ldots, n^p]$ and $X(r) = (1, r, \ldots, r^p)$.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

The limiting distributions of $\tilde{\gamma}$ and $\tilde{c}$ will be dependent on asymptotic behavior of the random variables $n^{-1}\sum_{t=1}^{n}\varphi'(\Delta y_t - \tilde{\gamma}'\Delta x_t)(y_{t-1} - \tilde{\gamma}'x_{t-1})$ and $n^{-1}\sum_{t=1}^{n}\varphi'(\Delta y_t - \tilde{\gamma}'\Delta x_t)\Delta_{\tilde{c}}x_t$. To derive the limiting distribution of $(\tilde{c},\tilde{\gamma})$, we assume that the following conditions hold.

Assumption B. $\tilde{c} = c + o_p(n^{1/4})$ and $n^{-1/2}D_n(\tilde{\gamma} - \gamma) = o_p(n^{1/4})$.

Assumptions similar to Assumption B are standard in the development of *M*-estimator asymptotics. It is related to Assumption (b) in Theorem 5.1 of Phillips (1995) and the assumption on $\tilde{\varepsilon}_t - \varepsilon_t$ in Theorem 1 of Lucas (1995). Notice that

$$\begin{aligned}\tilde{u}_t &= \Delta y_t - \tilde{c}(y_{t-1}/n) - \tilde{\gamma}'\Delta x_t + \tilde{c}\tilde{\gamma}'(x_{t-1}/n)\\ &= u_t - (\tilde{c} - c)(y_{t-1}^{s}/n) - (\tilde{\gamma} - \gamma)'\Delta_c x_t + (\tilde{c} - c)(\tilde{\gamma} - \gamma)'(x_{t-1}/n).\end{aligned}$$

Under Assumption B, $\tilde{u}_t - u_t$ satisfies the conditions in Lucas (1995) and Phillips (1995). The results of this paper can be obtained under different types of regularity conditions (for a discussion on consistency of *M*-estimators, see, e.g., Wooldridge, 1994). To cover a wide range of models, we simply assume that Assumption B holds. Denoting the limit of $n^{-1/2}D_n(\tilde{\gamma} - \gamma)$ by ξ_c and the limit of $\tilde{c}$ by η_c, the asymptotic distributions of the *M*-estimators $\tilde{c}$ and $\tilde{\gamma}$ are given in the following theorem.

THEOREM 1. *Given models (2.1)–(2.4), for all c in a compact set, under Assumptions A and B, the limiting distributions of nonlinear regression M-estimators* $\tilde{\gamma}$ *and* $\tilde{c}$ *are jointly determined by the following equations:*

$$\xi_c = \left[\int X_\eta(r)X_\eta(r)'dr\right]^{-1}\int X_\eta(r)d\tilde{S}_\eta(r),$$

$$\eta_c = \left[\int \underline{J}_{c\xi}(r)^2 dr\right]^{-1}\int \underline{J}_{c\xi}(r)d\tilde{S}_c(r),$$

where

$$X_\eta(r) = g(r) - \eta_c X(r), \qquad \underline{J}_{c\xi}(r) = J_c(r) - \xi_c' X(r),$$

$$\tilde{S}_c(r) = S_0(r) + \int_0^r \underline{J}_c(s)ds, \qquad \tilde{S}_\eta(r) = S_0(r) + (c - \eta_c)\int_0^r J_c(s)ds,$$

$$S_0(r) = \frac{\rho}{\delta}W_1(r) - \frac{\sqrt{\omega^2 - \rho^2}}{\delta}W_2(r), \qquad \underline{J}_c(r) = cJ_c(r) - \xi_c' g(r).$$

Remark 1. If φ is the true log density for u_t, we have $\rho = 1$ and $\omega^2 = \delta \geq 1$. The departure from Gaussianity in the data is completely determined by the parameter ω^2. When the data are generated by a Gaussian process, $\omega^2 = 1$ and the $W_2(\cdot)$ terms disappear from the limiting distribution. As ω^2 increases, the underlying distribution becomes more and more non-Gaussian.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Remark 2. In the stationary case, similar nonlinear regression estimators can be obtained from (1.5). However, under regularity conditions, closed-form solutions of the limiting distributions for these estimators can be derived, and it can be shown that they are first-order equivalent to the one-step Newton–Raphson estimators.

Remark 3. The maximum likelihood estimation based on this nonlinear regression generally provides a more efficient estimator than the OLS regression for the deterministic trend; this is also confirmed in the Monte Carlo experiment in Section 4. Because the true value of the local parameter c is unknown, this maximum likelihood estimator of the deterministic trend can not achieve the efficiency level that applies when the local parameter is known.

Remark 4. In practice, even if the exact distribution of the innovations is unknown, as long as the data have similar tail behavior as the density function used in the estimation, inferences based on these methods still have good sampling performance. Thus, we may consider adaptive (Hansen and Lee, 1994; Seo, 1996; Beelders, 1998) or partially adaptive (Bickel, 1982, p. 664; Potscher and Prucha, 1986; Xiao, 1999) estimation methods so that the data density can be approximated. For example, to capture the feature of heavy tails in economic and financial data, we may consider a partially adaptive estimator based on the Student-t distributions, which has wide applications in economic analysis. In case of t-distributed innovations, the log-likelihood is given by

$$L(\gamma, c) = \text{constant} + \frac{n}{2}\ln\Sigma - \frac{\nu+1}{2}\sum_{t=2}^{n}\ln\left\{1+\frac{\Sigma}{\nu}\left[\Delta y_t - \frac{cy_{t-1}}{n} - \gamma'\Delta x_t + \frac{c\gamma' x_{t-1}}{n}\right]^2\right\}.$$

Thus we consider a two-step partially adaptive estimator of (γ, c) in which the first step involves a preliminary estimation of the parameters ν and Σ. In the presence of general disturbance distributions, ν and Σ may lose their original meaning. However, when $\hat{\nu} \geq 0$ and $\hat{\Sigma} \geq 0$, $\hat{\nu}$ and $\hat{\Sigma}$ still can be interpreted as estimators of measures of the tail thickness and the spread of the disturbance distribution, and the partially adaptive estimator can still have good sampling properties.

Remark 5. From the proof of Theorem 1, it can be derived that the M-estimator for c has the following representation:

$$\tilde{c} = \frac{n^{-1}\sum_{t=1}^{n}\varphi'(\Delta y_t - \tilde{\gamma}'\Delta x_t)(y_{t-1} - \tilde{\gamma}' x_{t-1})}{n^{-2}\sum_{t=1}^{n}\varphi''(\Delta y_t - \tilde{\gamma}'\Delta x_t)(y_{t-1} - \tilde{\gamma}' x_{t-1})^2} + o_p(1).$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Remark 6. For simplicity we have assumed in this paper that u_t has unit variance. In the general case where the variance of u_t is σ^2, the analysis and results are parallel to those given in the paper, but the corresponding quantities $n^{-1/2}D_n(\tilde{\gamma} - \gamma)$, ω^2, and ρ need to be standardized by the scale parameter σ^2. If we modify Assumption A so that u_t are i.i.d. with mean zero and variance σ^2, then under the modified Assumption A, the partial sum process of u_t converges weakly to a Brownian motion with variance σ^2, i.e., $n^{-1/2}\sum_1^{[nr]} u_t \Rightarrow B_1(r) = \sigma W_1(r)$, where $W_1(r)$ is still defined as a standardized Brownian motion. Thus, $n^{-1/2}y^s_{[nr]}$ converges weakly to the corresponding Ornstein–Uhlenbeck process $J_c^*(r) = \int_0^r e^{c(r-s)}dB_1(s) = \sigma J_c(r)$, and $J_c(r) = \int_0^r e^{c(r-s)}dW_1(s)$. We standardize ω^2 and ρ by σ^2 as follows:

$$\omega^2_{\varphi.u} = \omega^2/\sigma^2, \rho_0 = \rho/\sigma^2$$

and deal with the correlation between $B_1(r)$ and $B_\varphi(r)$ explicitly by constructing random variable $v_t^* = \varphi'(u_t) + \rho_0 u_t$. It can be verified that v_t^* are i.i.d. with variance $\sigma^2(\omega^2_{\varphi.u} - \rho_0^2)$ and are uncorrelated with u_t. The partial sum process $n^{-1/2}\sum_1^{[nr]} v_t^*$ converges weakly to $\sigma\sqrt{\omega^2_{\varphi.u} - \rho_0^2}W_2(r)$, where $W_2(r)$ is a standard Brownian motion independent of $W_1(r)$. Thus $B_\varphi(r) = \sigma\{\sqrt{\omega^2_{\varphi.u} - \rho_0^2}W_2(r) - \rho_0 W_1(r)\}$. If we standardize $n^{-1/2}D_n(\tilde{\gamma} - \gamma)$ by σ^{-1} and denote the limiting variates of $n^{-1/2}\sigma^{-1}D_n(\tilde{\gamma} - \gamma)$ and $\tilde{c}$ by ξ_c and η_c, respectively, then, under the assumptions of Theorem 1 (with the modified Assumption A), the limiting distributions of the nonlinear regression M-estimators $\tilde{\gamma}$ and $\tilde{c}$ are jointly determined by the following equations:

$$\xi_c = \left[\int X_\eta(r)X_\eta(r)'dr\right]^{-1}\int X_\eta(r)d\tilde{S}^*_\eta(r),$$

$$\eta_c = \left[\int \underline{J}_{c\xi}(r)^2 dr\right]^{-1}\int \underline{J}_{c\xi}(r)d\tilde{S}^*_c(r),$$

where

$$X_\eta(r) = g(r) - \eta_c X(r), \qquad \underline{J}_{c\xi}(r) = J_c(r) - \xi_c' X(r),$$

$$\tilde{S}^*_c(r) = S_0^*(r) + \int_0^r \underline{J}_c(s)ds, \qquad \tilde{S}^*_\eta(r) = S_0^*(r) + (c - \eta_c)\int_0^r J_c(s)ds,$$

$$S_0^*(r) = \frac{\rho_0}{\delta}W_1(r) - \frac{\sqrt{\omega^2_{\varphi.u} - \rho_0^2}}{\delta}W_2(r), \qquad \underline{J}_c(r) = cJ_c(r) - \xi_c' g(r).$$

We can see that the only difference between the preceding results and those of Theorem 1 is that, in the general case, ω^2 and ρ^2 in Theorem 1 are replaced by their standardized versions $\omega^2_{\varphi.u}$ and ρ_0^2.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

If we take $\varphi(u) = -u^2$, the estimators are least squares regression estimators and Theorem 1 gives the results in Phillips and Xiao (1998). These estimators have likelihood interpretations when the process is actually Gaussian. However, even if the time series are not normal, the asymptotic results still hold. Denoting the nonlinear least squares regression estimators as $\tilde{\gamma}_{ls}$ and $\tilde{c}_{ls}$, we summarize the limiting distribution of the nonlinear least square estimators in the following corollary for convenience of later analysis.

COROLLARY 1. *Given models (2.1)–(2.4), for all c in a compact set, the limiting distributions of nonlinear least squares regression estimators* $\tilde{\gamma}_{ls}$ *and* $\tilde{c}_{ls}$ *are jointly determined by the following equations:*

$$\xi_c^* = \left[\int X_\eta(r)X_\eta(r)'dr\right]^{-1}\left[\int X_\eta(r)dV_\eta(r)\right],$$

$$\eta_c^* = \left[\int \underline{J}_{c\xi}(r)^2 dr\right]^{-1}\left[\int \underline{J}_{c\xi}(r)dU_{\xi}(r)\right],$$

where

$$X_c(r) = g(r) - cX(r), \qquad V_\eta(r) = W(r) - (\eta_c^* - c)\int_0^r J_c(s)ds,$$

$$U_{\xi}(r) = W(r) - \xi_c^{*\prime}\int_0^r X_c(s)ds,$$

and $W(r)$ *is a standard Brownian motion.*

3. HYPOTHESIS TESTS

3.1. Unit Root Tests against Local Alternatives Based on *M*-Estimators

This section considers unit root tests based on *M*-estimators proposed in the previous section. We are interested in the alternative hypothesis that α is less than unity. For alternatives that are distant from unity, the proposed tests will be consistent and will reject H_0 with probability close to one in large samples. Thus we are interested in local alternative hypothesis $H_1 : c < 0$.

In our discussion, we assume that φ is the true log density when likelihood ratio tests are considered. In other cases, φ can be a more general criterion function satisfying Assumption A. We start with the likelihood ratio test. Let $\hat{\gamma}$ be the restricted maximum likelihood estimator of γ under the null hypothesis of $c = 0$ and let $\tilde{c}$ and $\tilde{\gamma}$ be the unrestricted maximum likelihood estimators of c and γ analyzed in Section 2; then the likelihood ratio test for the null hypothesis of a unit root rejects H_0 for small values of $L(0, \hat{\gamma}) - L(\tilde{c}, \tilde{\gamma})$. The limiting distribution of the likelihood ratio statistic depends on the limiting distribu-

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

tions of both the unrestricted maximum likelihood estimator and the restricted maximum likelihood estimator. Under the null hypothesis, $n^{-1/2}y^s_{[nr]}$ converges weakly to $W_1(r)$, and, by Theorem 1, the limiting distributions of the unrestricted nonlinear regression estimators $\tilde{\gamma}$ and $\tilde{c}$ are jointly determined by the following equations:

$$\begin{cases} \eta_0 \int X_{\eta_0}(r)\underline{W}_{\xi_0}(r)dr = \int X_{\eta_0}(r)d\tilde{S}_0(r), \\ \eta_0 \int \underline{W}_{\xi_0}(r)^2 dr = \int \underline{W}_{\xi_0}(r)d\tilde{S}_0(r), \end{cases}$$

where

$$\underline{W}_{\xi_0}(r) = W_1(r) - \xi_0' X(r), \qquad \tilde{S}_0(r) = S_0(r) - \xi_0' X(r).$$

For the restricted estimator, notice that, under H_0, the log likelihood is simply $L(0,\gamma) = \sum\varphi(\Delta y_t - \gamma'\Delta x_t)$. The restricted MLE of γ satisfies the following first-order condition:

$$\sum_{t=1}^{n} \varphi'(\Delta y_t - \hat{\gamma}'\Delta x_t)\Delta x_t = 0.$$

When x_t contains a constant term, the corresponding element in Δx_t is zero, and, as a result, the regressor in the restricted model is actually of smaller dimension than x_t. To avoid the singularity problem in deriving the limiting distributions of $\hat{\gamma}$, we express this explicitly by rewriting the trend component in the restricted model as $\gamma'\Delta x_t = \beta'\check{x}_t$, where $\check{x}_t = Sx_t$ for some eliminator matrix S that eliminates redundant rows of x_t. Therefore $\check{x}_t$ is usually of smaller dimension than x_t, and the log likelihood can be rewritten as

$$\sum \varphi(\Delta y_t - \beta'\check{x}_t).$$

Assume that $G_n^{-1}\check{x}_{[nr]} \to \underline{X}(r)$ as $n \to \infty$, uniformly in $r \in [0,1]$. Then, under the null of a unit root, $\hat{\beta}$ has the following asymptotic distribution:

$$n^{1/2}G_n(\hat{\beta} - \beta) \Rightarrow \left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1} \int \underline{X}(r)dS_0(r). \tag{3.1}$$

Remark 7. If x_t does not contain a constant term, say, $x_t = t$, then $\check{x}_t$ is actually of the same dimension as x_t, and $\underline{X}(r) = g(r)$.

Remark 8. If φ is the log density of normal distribution, $S_0(r)$ is simply $W_1(r)$, and the limit distribution (3.1) reduces to the standard result of OLS detrending.

The asymptotic null distribution of the likelihood ratio statistic is summarized in Theorem 2.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

THEOREM 2. *Under Assumptions A and B and the null hypothesis* H_0: $c = 0$,

$$2[L(0,\hat{\gamma}) - L(\tilde{c},\tilde{\gamma})]$$
$$\Rightarrow \delta \int [\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)]^2 dr - 2\delta \int [\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)]\, dS_0(r)$$
$$+ \delta \int dS_0(r) \underline{X}(r)' \left[\int \underline{X}(r)\underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_0(r). \tag{3.2}$$

Remark 9. Likelihood ratio tests for the null hypothesis $c = c_0$ vs. the alternative $c \neq c_0$ can also be constructed. The principles of the proofs are the same, and the asymptotic results are similar to those of Theorem 2.

We may also construct test statistics directly based on the *M*-estimator of the unknown parameter. A natural candidate is simply the *M*-estimator of *c*. This can be treated as a generalization of the Dickey–Fuller (Dickey and Fuller, 1979) or Phillips–Perron (Phillips, 1987; Phillips and Perron, 1988) tests under *M*-estimation. From the result in Section 2, this is asymptotically equivalent to rejecting H_0 if

$$Z_M = \frac{n \sum_{t=1}^{n} \varphi'(\Delta y_t - \tilde{\gamma}' \Delta x_t)(y_{t-1} - \tilde{\gamma}' x_{t-1})}{\sum_{t=1}^{n} \varphi''(\Delta y_t - \tilde{\gamma}' \Delta x_t)(y_{t-1} - \tilde{\gamma}' x_{t-1})^2} < CV_\tau \tag{3.3}$$

for some critical value CV_τ. The asymptotic distribution is given in the following corollary.

COROLLARY 2. *Under Assumptions A and B and the null hypothesis, the statistic* Z_M *defined by (3.3) converges weakly to*

$$\left[\int \underline{W}_{\xi_0}(r)^2 dr \right]^{-1} \int \underline{W}_{\xi_0}(r) d\tilde{S}_0(r). \tag{3.4}$$

To obtain asymptotically valid tests for a unit root, we need to know the distributions given in (3.2) of Theorem 2 or (3.4) of Corollary 2. In the case that φ is the log density function, we can calculate the critical values by simulating W_1 and W_2. Notice that the departure from normality is characterized by ω^2, $\omega^2 = 1$ when the data is Gaussian. As ω^2 increases, the underlying distribution becomes more and more non-Gaussian. Thus, we may estimate and tabulate the asymptotic critical values for selected values of ω^2, say, $\omega^2 = 1,2,3,\ldots,10$, and so forth. For intermediate values of ω^2, critical values could be approximated by interpolation (for a discussion on related issues, see Hansen, 1995, p. 1155).

More generally, the distribution of u_t may not be known, so that ω^2, ρ, and δ must be estimated. The asymptotic null distribution is unaffected if the param-

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

eters are replaced by their consistent estimates. Thus, a robust estimate of the null distribution can be obtained by simulating the distribution with the unknown parameters replaced by their consistent estimates. Such robust tests could be inconvenient in practical analysis because the critical values will have to be calculated each time. An alternative way (see, e.g., Lucas, 1997) is to generate conservative critical values based on normal innovations.

Now we consider the power properties of the likelihood ratio statistic. The limiting distribution of (3.2) can be derived using the results of Theorem 1, and thus the power function can be obtained. We summarize the asymptotic results in the following theorem, which shows that the likelihood ratio test has nontrivial power against the local alternative.

THEOREM 3. *Under Assumptions A and B and the local alternative,*

$$
\begin{aligned}
&2[L(0,\hat{\gamma}) - L(\tilde{c},\tilde{\gamma})] \\
&\quad \Rightarrow \omega^2 \int [(\eta_c - c)\underline{J}_{c\xi}(r) + \xi_c' X_c(r)]^2 dr \\
&\qquad + 2\omega^2 \int [(\eta_c - c)\underline{J}_{c\xi}(r) + \xi_c' X_c(r)]\, dW_\varphi(r) + 2c\omega^2 \int J_c(r) dW_\varphi(r) \\
&\qquad + \omega^2 \int d\underline{S}_c(r)\underline{X}(r)' \left[\int \underline{X}(r)\underline{X}(r)' dr\right]^{-1} \int \underline{X}(r) d\underline{S}_c(r) \\
&\qquad - c^2\omega^2 \int J_c(r)^2 dr \\
&\qquad + 2c\omega^2 \int J_c(r)\underline{X}(r)' dr \left[\int \underline{X}(r)\underline{X}(r)' dr\right]^{-1} \int \underline{X}(r) d\underline{S}_c(r),
\end{aligned}
$$

where

$$
S_c(r) = S_0(r) + c\int_0^r J_c, \qquad \underline{S}_c(r) = S_0(r) - c\int_0^r J_c.
$$

Remark 10. Similar results can be obtained for the test Z_M.

Remark 11. Under the normality assumption, the likelihood ratio test can be constructed based on the least squares regression. It can be shown that, after dropping the asymptotically negligible terms, the likelihood ratio test rejects H_0 for small values of

$$
LR_c = \frac{\sum_{t=2}^{n} [\Delta y_t - \tilde{c}_{nl}(y_{t-1}/n) - \tilde{\gamma}_{nl}' \Delta x_t + \tilde{c}_{nl}\tilde{\gamma}_{nl}'(x_{t-1}/n)]^2}{\sum_{t=2}^{n} [\Delta y_t - \hat{\gamma}' \Delta x_t]^2},
$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

where the restricted maximum likelihood estimator $\hat{\gamma}$ is simply the least squares estimator of the following regression on the differenced data:

$$\Delta y_t = \gamma' \Delta x_t + \Delta y_t^s. \tag{3.5}$$

Using $\check{x}_t$ defined earlier in this section, (3.5) can be rewritten as $\Delta y_t = \beta' \check{x}_t + \Delta y_t^s$. The least squares estimator $\hat{\beta}_{ls}$ has the following asymptotic distribution:

$$n^{1/2} G(\hat{\beta}_{ls} - \beta) \Rightarrow \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dJ_c(r). \tag{3.6}$$

The asymptotic distributions of LR_c are given in the following theorem.

THEOREM 4.

(1) Under H_0,

$$\begin{aligned} n(LR_c - 1) &\Rightarrow \int [\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)]^2 dr - 2 \int [\eta_0 \underline{W}_{\xi_0}(r) + \xi_0' g(r)] \, dW(r)^2 dr \\ &\quad + \int dW(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dW(r). \end{aligned} \tag{3.7}$$

(2) Under H_c,

$$\begin{aligned} n(LR_c - 1) \Rightarrow \xi_c' \left\{ \int g(r) [X_\eta'(r) - (\eta_c + c) X'(r)] \, dr \right\} \xi_c + \int \tilde{R}_c(r)^2 dr \\ + \int dS_c(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_c(r) \\ - 2\xi_c' \int X_\eta(r) dW(r) - 2\eta_c \int J_c(r) d\underline{W}_\xi(r) \\ - c \int J_c(r) dW(r) - c \int J_c(r) dS_c(r), \end{aligned}$$

where

$$\tilde{R}_c(r) = \eta_c J_c(r) - (\eta_c + c) X'(r) \xi_c, \quad \textit{and} \quad \underline{W}_\xi(r) = W(r) - \xi_c' X(r).$$

Remark 12. In the special case that $\tilde{c}$ is the nonlinear least squares estimator, the limiting null distribution of Z_M is $[\int \underline{W}_{\xi_0}(r)^2 dr]^{-1} \int \underline{W}_{\xi_0}(r) d\underline{W}_{\xi_0}(r)$.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

3.2. Unit Root Test against a Point Alternative

Even for the simplest case where $x_t = 0$ (or d_t is known) and thus y_t^s is observable, there is no uniformly optimal estimator for c or uniformly optimal test for H_0. Under regularity conditions, the random variables $\sum_{t=1}^n \varphi'(\Delta y_t^s)(y_{t-1}^s/n)$ and $\sum_{t=1}^n \varphi''(\Delta y_t^s)(y_{t-1}^s/n)^2$ have a nondegenerate limiting distribution and are asymptotically jointly sufficient statistics for the local parameter c. Notice that the asymptotic sufficient statistic is two dimensional and we can not find a uniformly best estimate for c or a uniformly most powerful test for H_0 even asymptotically. Cox and Llatas (1991) studied the optimality of the maximum likelihood estimator for this case and showed that the optimal criterion function is a linear combination of the least squares score and the true score function, and the linear combination depends on the unknown parameter c. Because φ is the log density of u_t, asymptotic admissible tests could be constructed based on a linear combination of $\sum_{t=1}^n \varphi'(\Delta y_t^s)(y_{t-1}^s/n)$ and $\sum_{t=1}^n \varphi''(\Delta y_t^s)(y_{t-1}^s/n)^2$.

If we consider a unit root test against the simple point alternative $c = \bar{c} < 0$, then, in the case d_t is known, a most powerful test can be constructed based on the likelihood ratio statistic $L(0) - L(\bar{c})$ by the Neyman–Pearson lemma (e.g., King, 1988; Dufour and King, 1991). As a result, the asymptotic local power function can be calculated and a power envelope can be obtained as we change the values of $\bar{c}$ (Rothenberg and Stock, 1997).

When d_t is unknown, the trend coefficient γ has to be estimated to construct a feasible test. However, the use of an estimated γ changes the limiting distribution, and there is no most powerful test for the unit root hypothesis. In this case, a natural generalization of the Neyman–Pearson test for the null of $c = 0$ against the point alternative $c = \bar{c}$ is to reject for small values of the likelihood ratio $L(0, \hat{\gamma}) - L(\bar{c}, \bar{\gamma})$, where $\hat{\gamma}$ and $\bar{\gamma}$ are the maximum likelihood estimators for γ under the null and the alternative hypothesis, respectively. Elliott, Rothenberg, and Stock (1996) studied this test for the Gaussian case with a linear trend and constructed a power envelope based on such tests. In this section, we explore the asymptotic properties of such tests for cases with general disturbance distributions.

The asymptotic behavior of $L(0, \hat{\gamma})$ has been analyzed in Section 3.1. For $\bar{\gamma}$, it can be shown that

$$n^{-1/2} D_n(\bar{\gamma} - \gamma) \Rightarrow \left[\int X_{\bar{c}}(r) X_{\bar{c}}(r)' dr \right]^{-1} \int X_{\bar{c}}(r) d\bar{S}_c(r), \tag{3.8}$$

where

$$\bar{S}_c(r) = S_0(r) - (\bar{c} - c) \int_0^r J_c, \qquad X_{\bar{c}}(r) = g(r) - \bar{c} X(r).$$

By an asymptotic expansion and using the results of (3.1) and (3.8), the asymptotic distribution of the likelihood ratio statistic can be derived.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

THEOREM 5. *For all c in a compact set,*

$$
\begin{aligned}
&2[L(0,\hat{\gamma}) - L(\bar{c},\bar{\gamma})] \\
&\quad \Rightarrow \delta \int \left[(\bar{c}-c) J_c(r) + X_{\bar{c}}(r)' \left[\int X_{\bar{c}}(r) X_{\bar{c}}(r)' dr \right]^{-1} \int X_{\bar{c}}(r) d\bar{S}_c(r) \right]^2 dr \\
&\qquad - 2\delta \int \left[(\bar{c}-c) J_c(r) + X_{\bar{c}}(r) \left[\int X_{\bar{c}}(r) X_{\bar{c}}(r)' dr \right]^{-1} \int X_{\bar{c}}(r) d\bar{S}_c(r) \right] dS_0(r) \\
&\qquad + \delta \int d\underline{S}_c(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_c(r) - c^2 \delta \int J_c(r)^2 dr \\
&\qquad - 2c\delta \int J_c(r) dS_0(r) + 2c\delta \int J_c(r) \underline{X}(r)' dr \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_c(r).
\end{aligned}
$$

Under the null of $c = 0$ and the point alternative $c = \bar{c}$, we obtain the following limiting distributions.

COROLLARY 3.

(1) Under $c = 0$,

$$
\begin{aligned}
&2[L(0,\hat{\gamma}) - L(\bar{c},\bar{\gamma})] \\
&\quad \Rightarrow \delta \int \left[\bar{c} W_1(r) + X_{\bar{c}}(r)' \left[\int X_{\bar{c}}(r) X_{\bar{c}}(r)' dr \right]^{-1} \int X_{\bar{c}}(r) d\bar{S}_0(r) \right]^2 dr \\
&\qquad - 2\delta \int \left[\bar{c} W_1(r) + X_{\bar{c}}(r)' \left[\int X_{\bar{c}}(r) X_{\bar{c}}(r)' dr \right]^{-1} \int X_{\bar{c}}(r) d\bar{S}_0(r) \right] dS_0(r) \\
&\qquad + \delta \int dS_0(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_0(r).
\end{aligned}
$$

(2) Under $c = \bar{c}$,

$$
\begin{aligned}
&2[L(0,\hat{\gamma}) - L(\bar{c},\bar{\gamma})] \\
&\quad \Rightarrow -\delta \int dS_0(r) X_{\bar{c}}(r)' \left[\int X_{\bar{c}}(r) X_{\bar{c}}(r)' dr \right]^{-1} \int X_{\bar{c}}(r) dS_0(r) \\
&\qquad + \delta \int d\underline{S}_{\bar{c}}(r) \underline{X}(r)' \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \int \underline{X}(r) dS_{\bar{c}}(r) - \bar{c}^2 \delta \int J_{\bar{c}}(r)^2 dr \\
&\qquad - 2\bar{c}\delta \int J_{\bar{c}}(r) dS_0(r) + 2\bar{c}\delta \int J_{\bar{c}}(r) \underline{X}(r)' dr \left[\int \underline{X}(r) \underline{X}(r)' dr \right]^{-1} \\
&\qquad \times \int \underline{X}(r) dS_{\bar{c}}(r). \qquad \textbf{(3.9)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Remark 13. In the special case that φ is the log density of normal distribution, the likelihood ratio test rejects H_0 for small values of

$$LR_c = \frac{\sum_{t=2}^{n} [\Delta y_t - \bar{c}(y_{t-1}/n) - \bar{\gamma}'\Delta x_t + \bar{c}\bar{\gamma}'(x_{t-1}/n)]^2}{\sum_{t=2}^{n} [\Delta y_t - \hat{\gamma}'\Delta x_t]^2}. \tag{3.10}$$

The limiting distributions of (3.10) can be derived as a corollary of Theorem 5. In particular, when x_t is a linear trend, the results reduce to the distribution given by Elliott et al. (1996). They also derive the power envelope based on distribution (3.9) for the case where φ is log normal and x_t is a linear trend. Obviously, in the case of non-Gaussian innovations, this distribution depends on the parameter ω^2. Given the value of ω^2, the power envelope can be derived similarly. Monte Carlo evidence indicates that substantial power increase occurs as the parameter ω^2 increases. This suggests a potential efficiency gain from using the distributional information in the unit root tests.

3.3. QD *M*-Detrended Unit Root Tests

The *M*-estimation can be coupled with quasi-differencing (QD) to construct a nearly efficient *M*-detrended unit root test. Again, like the tests considered in the previous sections, different types of unit root tests based on the efficient *M*-detrended data can be constructed. We analyze the coefficient based test in this section; other tests can be analyzed in a similar way.

For some appropriate choice $\bar{c}$, we calculate the maximum likelihood estimators for γ under the hypothesis $c = \bar{c}$,

$$\bar{\gamma} = \arg\max \sum_{t=1}^{n} \varphi(\Delta_{\bar{c}} y_t - \gamma'\Delta_{\bar{c}} x_t), \tag{3.11}$$

and construct the detrended y_t based on $\bar{\gamma}$,

$$\bar{y}_t^s = y_t - \bar{\gamma}' x_t.$$

Reestimating c based on the *M*-estimator of the autoregressive coefficient of $\bar{y}_t^s$, i.e.,

$$\check{c} = \arg\max \sum_{t=1}^{n} \varphi(\Delta y_t - c(y_{t-1}/n) - \bar{\gamma}'\Delta x_t + c\bar{\gamma}'(x_{t-1}/n)),$$

the efficient *M*-detrended unit root test can then be constructed based on $\bar{Z}_M = \check{c}$.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Notice that the partial sum process based on $\bar{y}_t^s$ has the following asymptotic behavior:

$$\frac{1}{\sqrt{n}}\bar{y}^s_{[nr]} \Rightarrow J_c(r) - X(r)'\left[\int X_{\bar{c}}X'_{\bar{c}}\right]^{-1}\int X_{\bar{c}}\,d\bar{S}_c := \bar{J}_{\bar{c}}(r).$$

We can derive the limiting null distribution and the power function of $\bar{Z}_M = \check{c}$.

THEOREM 6.

(1) Under $c = 0$,

$$\check{c} \Rightarrow \left[\int \underline{W}^2_{\bar{c}}\right]^{-1}\left[\int \underline{W}_{\bar{c}}\,dS_0 - \int \underline{W}_{\bar{c}}X'_c\left[\int X_{\bar{c}}X'_{\bar{c}}\right]^{-1}\int X_{\bar{c}}\,d\bar{S}_0\right]. \tag{3.12}$$

(2) Under the alternative hypothesis,

$$\check{c} \Rightarrow c + \left[\int \bar{J}^2_{\bar{c}}\right]^{-1}\left[\int \bar{J}_{\bar{c}}\,dS_0 - \int \bar{J}_{\bar{c}}X'_c\left[\int X_{\bar{c}}X'_{\bar{c}}\right]^{-1}\int X_{\bar{c}}\,d\bar{S}_0\right].$$

Remark 14. In the case that $\varphi(u) = -u^2$, it can be easily verified that under the null

$$\check{c} \Rightarrow \left[\int \underline{W}^2_{\bar{c}}\right]^{-1}\int \underline{W}_{\bar{c}}\,d\underline{W}_{\bar{c}},$$

which is exactly the limit distribution of the quasi-differencing detrended Phillips Z_α test (see, inter alia, Phillips and Xiao, 1998). If we choose x_t to be a constant term or a linear trend, we obtain the limiting result of Elliott et al. (1996).

Remark 15. For time series with general serially correlated residuals, a nonparametrically modified estimator, say, $\hat{c}^+$, can be used, and the same limiting results follow.

4. MONTE CARLO RESULTS

We conducted some limited Monte Carlo experiments to examine the sampling performance of the nonlinear joint estimation of the trend coefficients and the local parameter and testing procedures based on them. In particular, we compared the finite sample performance of the nonlinear joint estimator of the deterministic trend coefficient with other conventional estimators and compared the power properties of unit root tests based on different detrending procedures. The model used for data generation was the following model:

$$(DGP)\begin{cases} y_t = \gamma' x_t + y_t^s, \\ y_t^s = \alpha y_{t-1}^s + u_t, \qquad t = 1,\ldots,n, \end{cases}$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

where the true value of γ is 0 and $\{u_t\}$ is an i.i.d. sequence of t-distributions with three degrees of freedom. We standardized u_t so that it has unity variance. Two sample sizes were considered: $n = 100$, $n = 200$. The number of iterations is 2,000 in each case, and the initial value of y^s is set at 0.

We first examined the estimation of deterministic trends, i.e., γ. We considered the leading case of a linear time trend, i.e., $x_t = (1, t)$. Notice that because the intercept term is not consistently estimable, we focused our attention on the estimation of the coefficient of t. The following estimators of the deterministic trend coefficient were compared:

(1) Ordinary least squares estimator of the trend coefficient, denoted as OLS.
(2) Least squares estimator of the trend coefficient based on the quasi-differenced data, denoted as QDLS.
(3) M-estimator of the trend coefficient based on the quasi-differenced data, i.e., $\bar{\gamma}$ in (3.11), denoted as QDM.
(4) Maximum likelihood estimator of the trend coefficient based on the nonlinear regression, i.e., $\tilde{\gamma}$ in (2.6), denoted as NLM.

We considered different data sets generated by (DGP) with $\alpha = 1$, 0.95, 0.9, 0.85, and 0.8 and sample sizes of 100 and 200. The prespecified $\bar{c}$ in quasi-differencing is -10 for the case $n = 100$ and -20 for the $n = 200$ case (other choices of prespecified $\bar{c}$ were also tried). Table 1 reports the estimation bias for the four estimators in different cases, and Table 2 reports the variances of these estimators. Notice that these estimators are unbiased and the mean squared errors are dominated by the variances. We also depicted the (kernel smoothed) simulation densities of these estimators corresponding to different data sets. The information about these graphics is given here:

TABLE 1. Estimation of the deterministic trend (bias)

	OLS	QDLS	QDM	NLM
		$n = 100$		
$\alpha = 1$	−0.0010215	−0.0006520	−0.00100501	−0.0010499
$\alpha = 0.95$	0.0003870	0.0003634	0.00010376	0.0000627
$\alpha = 0.90$	0.0004275	0.0003949	−0.00000908	−0.0000471
$\alpha = 0.85$	0.0003171	0.0002694	−0.00006837	−0.0000662
$\alpha = 0.80$	0.0002477	0.0001965	−0.00007931	−0.0000246
		$n = 200$		
$\alpha = 1$	0.0004712	0.0004742	−0.0015339	−0.0007481
$\alpha = 0.95$	0.0001856	0.0001727	0.0000656	−0.000102
$\alpha = 0.90$	0.0000221	0.0000161	0.0000122	0.00000251
$\alpha = 0.85$	0.0000179	0.0000153	−0.0000459	−0.0000664
$\alpha = 0.80$	0.0001281	0.0000954	−0.0000559	−0.0000325

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

TABLE 2. Estimation of the deterministic trend (variance)

	OLS	QDLS	QDM	NLM
		$n = 100$		
$\alpha = 1$	0.1158501	0.1104706	0.1056588	0.0090731
$\alpha = 0.95$	0.0295969	0.0271542	0.0202029	0.0189686
$\alpha = 0.90$	0.0159903	0.0141669	0.0100533	0.0106364
$\alpha = 0.85$	0.0108151	0.0099818	0.0066257	0.0070036
$\alpha = 0.80$	0.0081582	0.0078926	0.0054088	0.0052918
		$n = 200$		
$\alpha = 1$	0.0763741	0.0751235	0.0731166	0.0618703
$\alpha = 0.95$	0.0145821	0.0133685	0.0096584	0.0092456
$\alpha = 0.90$	0.0057442	0.0055467	0.0036155	0.0038017
$\alpha = 0.85$	0.0038246	0.0035625	0.0021288	0.0023156
$\alpha = 0.80$	0.0024754	0.0022922	0.0017279	0.0016296

(1) Figure 1: simulation densities of these estimators when the sample size is 100 and the true $\alpha = 1$.
(2) Figure 2: simulation densities when $n = 100$ and the true α is 0.9.
(3) Figure 3: simulation densities when $n = 200$ and the true $\alpha = 1$.
(4) Figure 4: simulation densities when $n = 200$ and the true α is 0.9.

Some general conclusions can be drawn from these Monte Carlo results. The estimators QDM and NLM, which use the distributional information, have much

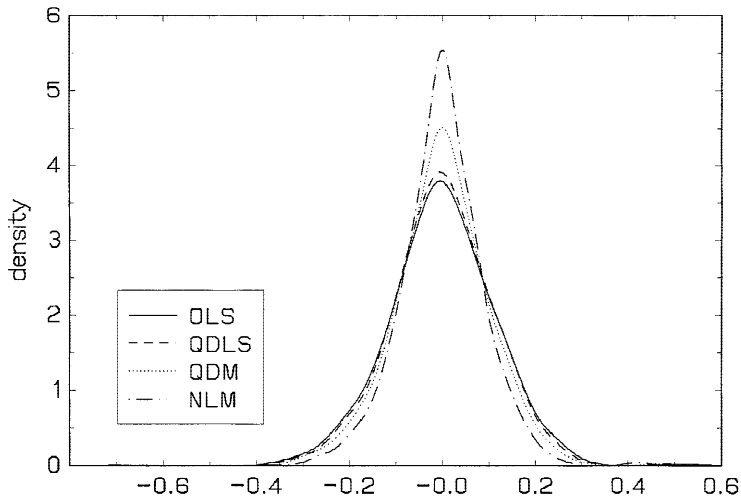


FIGURE 1. Simulation densities, $n = 100$, $\alpha = 1$.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

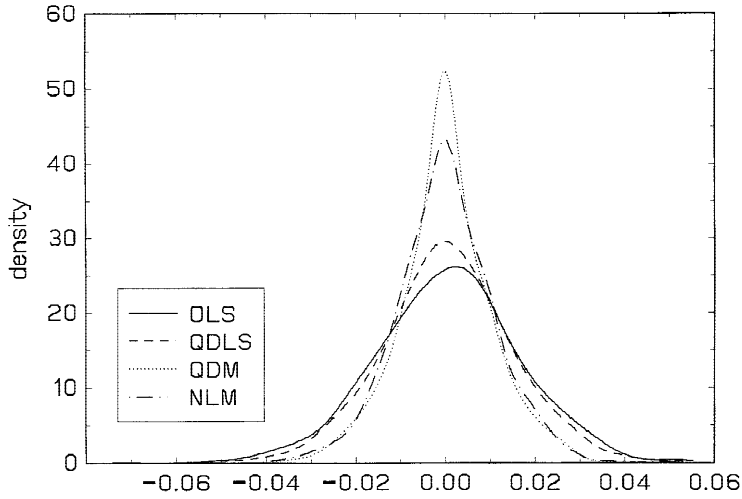


FIGURE 2. Simulation densities, $n = 100$, $\alpha = 0.9$.

better sampling performance than OLS and QDLS. Comparing QDM and NLM, we can treat NLM as the *M*-estimator using a nonlinearly estimated c, but QDM uses a prespecified c. (The one-step estimator was also examined and was found to be worse than the nonlinear estimator.) The Monte Carlo results indicate that the nonlinear joint estimator $\tilde{\gamma}$ (using a jointly estimated c in quasi-differencing)

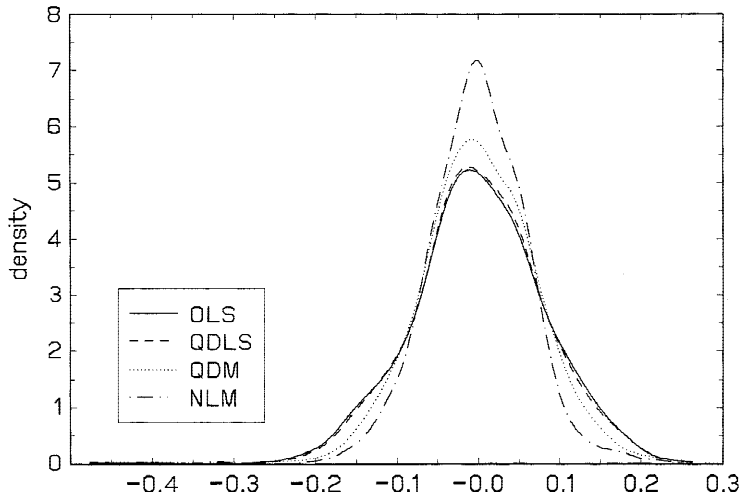


FIGURE 3. Simulation densities, $n = 200$, $\alpha = 1$.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

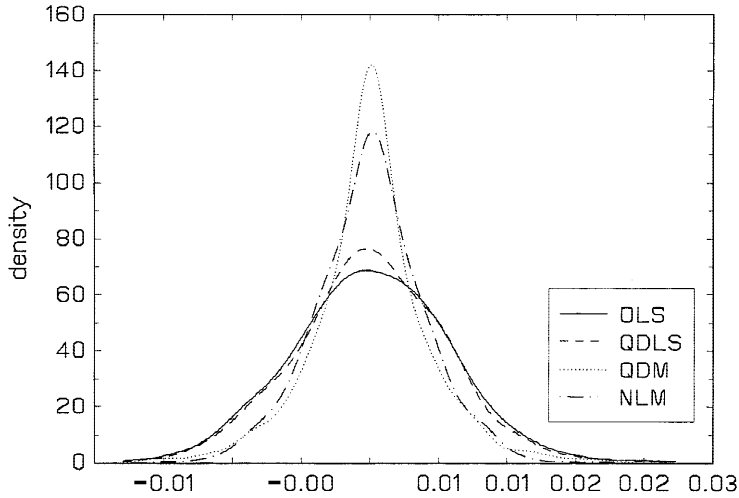


FIGURE 4. Simulation densities, $n = 200$, $\alpha = 0.9$.

has in general pretty good sampling performance and has avoided the additional issue of choosing a prespecified $\bar{c}$. As we can anticipate, when the true c value is close to the prespecified $\bar{c}$, QDM has slightly better performance over NLM. However, in other cases, NLM gives relatively better results. The evidence is quite clear in the four figures.

We also examined the effect of these estimation procedures on the power properties of unit root tests. In particular, we considered the coefficient-based unit root tests using these four different detrending procedures and the likelihood ratio test (3.2) based on the nonlinear joint estimation. Thus, the five tests examined in the Monte Carlo are

(1) the Dickey–Fuller coefficient test based on OLS detrending, denoted by OLS;
(2) the QD detrended DF test (based on least square regression on the quasi-differenced data), denoted by QD;
(3) the $\bar{Z}_M$ (3.12) test based on M-estimation plus quasi-differencing, denoted by QDM;
(4) the Z_M (3.3) test based on the nonlinear M-estimation, denoted by NL;
(5) the likelihood ratio test (3.2) based on the nonlinear M-estimation, denoted by LR.

To provide a power comparison among the different tests, size-corrected power is reported (for discussions on the use of size-corrected power, also see Stock, 1995; Cheung and Lai, 1997). The corresponding critical values were calculated from a direct simulation using 20,000 replications. Both the demeaned test and the detrended test are examined. In particular, Table 3 reports the empirical power of the demeaned tests, and Table 4 reports the power of the tests when a linear time trend is removed. Figure 5 depicts the power functions of

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

TABLE 3. Size-adjusted empirical power (demeaned case)

	OLS	QD	QDM	NL	LR
$\alpha = 0.975$	0.0976	0.1002	0.1260	0.1266	0.1362
$\alpha = 0.95$	0.1792	0.1822	0.2544	0.2540	0.2970
$\alpha = 0.925$	0.2982	0.3076	0.4682	0.4660	0.4900
$\alpha = 0.90$	0.4546	0.4688	0.7086	0.7082	0.6676
$\alpha = 0.875$	0.6278	0.6416	0.8896	0.8898	0.8101
$\alpha = 0.85$	0.7874	0.7962	0.9672	0.9666	0.8918
$\alpha = 0.825$	0.8968	0.9016	0.9908	0.9912	0.9502
$\alpha = 0.80$	0.9578	0.9606	0.9980	0.9982	0.9799

the demeaned tests, and Figure 6 depicts those of the detrended tests. A general conclusion that can be drawn immediately from the simulation results is that the testing procedures using distributional information have substantially improved power properties. These Monte Carlo results indicate that the empirical power functions of tests QDM, NL, and LR are well above the power functions of least-square-based tests. Tests based on the joint estimation have reasonably good performance.

5. CONCLUDING REMARKS

We studied likelihood-based estimation and tests in a nonstationary autoregressive time series model with unknown deterministic trends and general disturbance distributions. In particular, a joint estimation based on a nonlinear regression was studied. Asymptotic analysis on the *M*-estimators and related testing procedures were presented. The finite sample performance of these estimators and testing procedures was examined in Monte Carlo experiments. In

TABLE 4. Size-adjusted empirical power (detrended case)

	OLS	QD	QDM	NL	LR
$\alpha = 0.975$	0.0624	0.0638	0.0786	0.0778	0.1266
$\alpha = 0.95$	0.1048	0.1068	0.1592	0.1584	0.2654
$\alpha = 0.925$	0.1936	0.1950	0.3166	0.3160	0.4462
$\alpha = 0.90$	0.3192	0.3222	0.5434	0.5406	0.6228
$\alpha = 0.875$	0.4804	0.4822	0.7684	0.7690	0.7776
$\alpha = 0.85$	0.6474	0.6522	0.9134	0.9126	0.8728
$\alpha = 0.825$	0.7974	0.8024	0.9736	0.9726	0.9336
$\alpha = 0.80$	0.8982	0.9006	0.9926	0.9918	0.9676

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

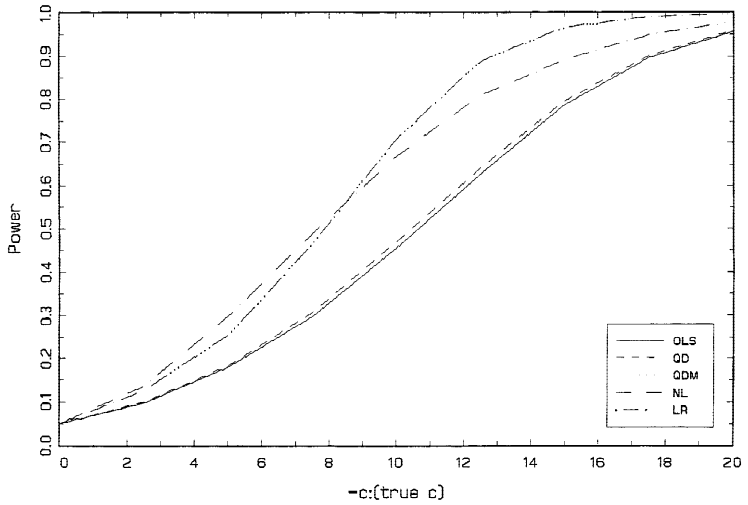


FIGURE 5. Size corrected power of unit root tests (demeaned case).

practice, even if the exact distribution of the innovations is unknown, if the data have similar tail behavior as the density function used in the estimation, inference based on these methods should have good sampling properties.

Our analysis may be extended in different directions of possible econometric interest. The approach of this paper readily extends to *M*-estimation with gen-

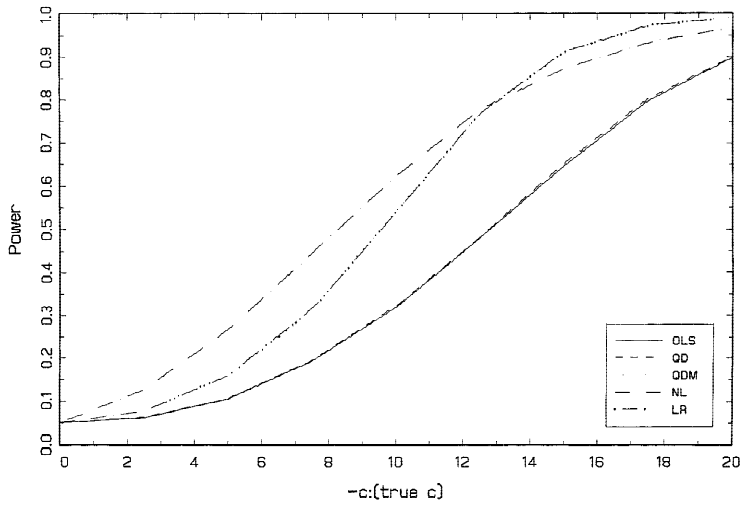


FIGURE 6. Size corrected power of unit root tests (linear time trend).

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

erally serial correlated innovations, in which case the one-sided long-run variance enters the limiting distribution; Structural breaks in the deterministic trend may be incorporated into the model. Adaptive or partially adaptive methods using approximations of the data densities may be investigated. It is also of considerable interest to extend this analysis to vector autoregressive time series with cointegrated variables.

REFERENCES

Beelders, O. (1998) Adaptive Estimation and Unit Root Tests. Working paper, Emory University.

Bickel, P. (1982) On adaptive estimation. *Annals of Statistics* 10, 647–671.

Chan, N.H. & C.Z Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063. Also appeared in *Forecasting and Modeling in Statistics and Econometrics* (Springer–Verlag, 1996).

Cheung, Y.W. & K.S. Lai (1997) Bandwidth selection, prewhitening, and the power of the Phillips–Perron test. *Econometric Theory* 13, 679–691.

Cox, D. & I. Llatas (1991) Maximum likelihood type estimation for nearly non-stationary autoregressive time series. *Annals of Statistics* 19, 1109–1128.

Dickey, D.A. & W.A. Fuller (1979) Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Dufour, J. & M.L. King (1991) Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary and nonstationary AR(1) errors. *Journal of Econometrics* 47, 115–143.

Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Hansen, B. (1995) Rethinking the univariate approach to unit root tests: How to use covariates to increase power. *Econometric Theory* 11, 1148-1171.

Hansen, B. & S. Lee (1994) Asymptotic theory for GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29–52.

Juhl, T. (1999) Testing for Cointegration Using M Estimators. Working paper, University of Kansas.

Juhl, T. & Z. Xiao (2000) Power Functions and Envelopes for Unit Root Tests. Preprint.

King, M.L. (1988) Towards a theory of point optimal testing. *Econometric Reviews* 6, 169–218.

Knight, K. (1991) Limit theory for M estimators in an integrated infinite variance process. *Econometric Theory* 7, 200–212.

Lucas, A. (1994) An outlier robust unit root test with an application to the extended Nelson–Plosser data. *Journal of Econometrics* 66, 153–174.

Lucas, A. (1995) Unit root tests based on M estimators. *Econometric Theory* 11, 331–346.

Lucas, A. (1997) Cointegration testing using pseudolikelihood ratio tests. *Econometric Theory* 13, 149–169.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated process: Part I. *Econometric Theory* 4, 468–497.

Park, J.Y. & P.C.B. Phillips (1989) Statistical inference in regressions with integrated process: Part II. *Econometric Theory* 5, 95–131.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory of autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1044.

Phillips, P.C.B. (1995) Robust nonstationary regression, *Econometric Theory* 12, 912–951.

Phillips, P.C.B. & P. Perron (1988) Testing for unit roots in time series regression. *Biometrika* 75, 335–346.

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Phillips, P.C.B. & Z. Xiao (1998) A primer on unit root testing. *Journal of Economic Survey* 12, 423–469.
Potscher, B. & I. Prucha (1986) A class of partially adaptive one-step M estimators for the nonlinear regression model with dependent observations. *Journal of Econometrics* 32, 219–251.
Rothenberg, T. & J. Stock (1997) Inference in a nearly integrated autoregressive model with nonnormal innovations. *Journal of Econometrics* 80, 269–286.
Schmidt, P. & P.C.B. Phillips (1992) Testing for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–288.
Seo, B. (1996) *Essays on Time Series Econometrics*, Ph.D. Dissertation, University of Rochester.
Stock, J.H. (1995) Unit roots, structural breaks, and trends. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2739–2841. Amsterdam: North-Holland.
Wooldridge, J.M. (1994) Estimation and inference for dependent processes. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: North-Holland.
Xiao, Z. (1999) Unit Root Tests Based on Partially Adaptive Estimation. Working paper, University of Illinois.

APPENDIX: PROOFS

Proof of Theorem 1. By definition, the estimators $(\tilde{c}, \tilde{\gamma})$ solve the following equation system:

$$\begin{cases} \text{(i)}: \sum_{t=1}^{n} \varphi'(\Delta y_t - \tilde{c}(y_{t-1}/n) - \tilde{\gamma}'\Delta x_t + \tilde{c}\gamma'(x_{t-1}/n))\Delta_{\tilde{c}} x_t = 0, \\ \text{(ii)}: \sum_{t=1}^{n} \varphi'(\Delta y_t - \tilde{c}(y_{t-1}/n) - \tilde{\gamma}'\Delta x_t + \tilde{c}\gamma'(x_{t-1}/n))\tilde{y}^s_{t-1} = 0, \end{cases}$$

where $\tilde{y}^s_{t-1} = y_{t-1} - \tilde{\gamma}' x_{t-1}$. For simplicity, we denote that

$$\tilde{u}_t = \Delta y_t - \tilde{c}(y_{t-1}/n) - \tilde{\gamma}'\Delta x_t + \tilde{c}\tilde{\gamma}'(x_{t-1}/n);$$

then

$$\tilde{u}_t = u_t - (\tilde{c} - c)(y^s_{t-1}/n) - (\tilde{\gamma} - \gamma)'\Delta_c x_t + (\tilde{c} - c)(\tilde{\gamma} - \gamma)'(x_{t-1}/n).$$

Under Assumption B, for all c in a compact set, the left-hand side of equation (i) is asymptotically equivalent to

$$\text{(i}'\text{)}: \sum_{t=1}^{n} \varphi'(u_t)\Delta_{\tilde{c}} x_t - (\tilde{c} - c)\sum_{t=1}^{n} \varphi''(u_t)\left(\frac{y^s_{t-1}}{n}\right)\Delta_{\tilde{c}} x_t$$

$$- (\tilde{\gamma} - \gamma)\sum_{t=1}^{n} \varphi''(u_t)\Delta_c x_t' \Delta_{\tilde{c}} x_t + (\tilde{c} - c)(\tilde{\gamma} - \gamma)\sum_{t=1}^{n} \varphi''(u_t)\left(\frac{x'_{t-1}}{n}\right)\Delta_{\tilde{c}} x_t .$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Under the assumptions of Theorem 1, the following asymptotics hold:

$$F_n^{-1}\Delta_{\tilde{c}}x_{[nr]} \Rightarrow g(r) - \eta_c X(r) = X_\eta(r),$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varphi'(u_t) \Rightarrow B_\varphi(r) = \omega W_\varphi(r) = BM(\omega^2),$$

$$\frac{1}{\sqrt{n}}\tilde{y}^s_{[nr]} \Rightarrow J_c(r) - \xi_c' X(r) = \underline{J}_{c\xi}(r),$$

where $J_c(r) = \int_0^r e^{c(r-s)}dW_1(s)$ and $W_1(r)$ is the weak limit of $n^{-1/2}\sum_1^{[nr]}u_t$. As a result, it can be verified that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varphi'(u_t)\Delta_{\tilde{c}}x_t F_n^{-1} \Rightarrow \int_0^1 dB_\varphi X_\eta, \tag{A.1}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varphi''(u_t)(\tilde{c}-c)\left(\frac{y^s_{t-1}}{n}\right)\Delta_{\tilde{c}}x_t F_n^{-1} \Rightarrow -\delta(\eta_c - c)\int J_c X_\eta, \tag{A.2}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varphi''(u_t)(\tilde{\gamma}-\gamma)'\Delta_c x_t'\Delta_{\tilde{c}}x_t F_n^{-1} \Rightarrow -\delta\xi_c'\int X_c X_\eta, \tag{A.3}$$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varphi''(u_t)(\tilde{c}-c)(\tilde{\gamma}-\gamma)'\left(\frac{x_{t-1}}{n}\right)\Delta_{\tilde{c}}x_t F_n^{-1} \Rightarrow -\delta(\eta_c - c)\xi_c'\int XX_\eta. \tag{A.4}$$

Notice that

$$\varphi'(u_t) = [\varphi'(u_t) + \rho u_t] - \rho u_t = v_t - \rho u_t.$$

Thus

$$B_\varphi(r) = \sqrt{\omega^2 - \rho^2}\,W_2(r) - \rho W_1(r). \tag{A.5}$$

Denote

$$\tilde{S}_c(r) = S_0(r) + \int_0^r \underline{J}_c(s)ds, \qquad \tilde{S}_\eta(r) = S_0(r) + (c - \eta_c)\int_0^r J_c(s)ds,$$

$$S_0(r) = \frac{\rho}{\delta}W_1(r) - \frac{\sqrt{\omega^2-\rho^2}}{\delta}W_2(r),\ \underline{J}_c(r) = cJ_c(r) - \xi_c' g(r);$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

then, by the results of (A.1)–(A.5) and (i′), we have

$$\xi_c = \left[\int X_\eta(r) X_\eta(r)' dr \right]^{-1} \int X_\eta(r) d\tilde{S}_\eta(r).$$

Similarly, under the given assumptions, the left-hand side of equation (ii) is asymptotically equivalent to

$$\text{(ii}'\text{)}: \sum_{t=1}^{n} \varphi'(u_t)\tilde{y}^s_{t-1} - (\tilde{c} - c) \sum_{t=1}^{n} \varphi''(u_t) \left(\frac{y^s_{t-1}}{n} \right) \tilde{y}^s_{t-1}$$

$$- (\tilde{\gamma} - \gamma)' \sum_{t=1}^{n} \varphi''(u_t) \Delta_c x_t \tilde{y}^s_{t-1} + (\tilde{c} - c)(\tilde{\gamma} - \gamma)' \sum_{t=1}^{n} \varphi''(u_t) \left(\frac{x_{t-1}}{n} \right) \tilde{y}^s_{t-1},$$

and by calculations of weak limits of the sample covariances $(1/n)\sum_{t=1}^{n} \varphi'(u_t)\tilde{y}^s_{t-1}$, $(1/n^2)\sum_{t=1}^{n} \varphi''(u_t) y^s_{t-1}\tilde{y}^s_{t-1}$, $(1/n)(\tilde{\gamma} - \gamma)' \sum_{t=1}^{n} \varphi''(u_t)\Delta_c x_t \tilde{y}^s_{t-1}$, and $(1/n)(\tilde{\gamma} - \gamma)' \times \sum_{t=1}^{n} \varphi''(u_t)(x_{t-1}/n)\tilde{y}^s_{t-1}$, following a similar argument as the previous proof, we get

$$\eta_c \int \underline{J}_{c\xi}(r)^2 dr = \int \underline{J}_{c\xi}(r) d\tilde{S}_c(r). \qquad \blacksquare$$

Proof of Theorem 2. Notice that the likelihood ratio statistic is

$$L(0, \hat{\gamma}) - L(\tilde{c}, \tilde{\gamma})$$

$$= \sum_t \varphi(\Delta y_t - \hat{\beta}' \breve{x}_t) - \sum_t \varphi(\Delta y_t - \tilde{c}(y_{t-1}/n) - \tilde{\gamma}' \Delta x_t + \tilde{c}\gamma'(x_{t-1}/n)).$$

By an asymptotic expansion, it can be shown that, under H_0 and the given assumptions, the likelihood ratio statistic has the following approximation:

$$\frac{\tilde{c}}{n} \sum_{t=1}^{n} \varphi'(u_t)(y_{t-1} - \tilde{\gamma}' x_{t-1}) - \frac{\tilde{c}}{n} \sum_{t=1}^{n} \varphi''(u_t)(y_{t-1} - \tilde{\gamma}' x_{t-1})(\tilde{\gamma} - \gamma)' \Delta x_t$$

$$- \frac{\tilde{c}^2}{2n^2} \sum_{t=1}^{n} \varphi''(u_t)(y_{t-1} - \tilde{\gamma}' x_{t-1})^2 + \sum_{t=1}^{n} \varphi'(u_t)(\tilde{\gamma} - \gamma)' \Delta x_{t-1}$$

$$- \sum_{t=1}^{n} \varphi'(u_t)(\hat{\beta} - \beta)' \breve{x}_t - \frac{1}{2} \sum_{t=1}^{n} \varphi''(u_t)[(\tilde{\gamma} - \gamma)' \Delta x_t]^2$$

$$+ \frac{1}{2} \sum_{t=1}^{n} \varphi''(u_t)[(\hat{\beta} - \beta)' \breve{x}_t]^2 + o_p(1). \qquad \textbf{(A.6)}$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

It can be derived that the components in (A.6) have the following limits:

$$\frac{\tilde{c}}{n}\sum_{t=1}^{n}\varphi'(u_t)(y_{t-1}-\tilde{\gamma}'x_{t-1}) \Rightarrow \eta_0\int \underline{W}_{\xi_0}(r)dB_\varphi(r),$$

$$\frac{\tilde{c}}{n}\sum_{t=1}^{n}\varphi''(u_t)(y_{t-1}-\tilde{\gamma}'x_{t-1})(\tilde{\gamma}-\gamma)'\Delta x_t \Rightarrow -\delta\eta_0\xi_0'\int g(r)\underline{W}_{\xi_0}(r)dr,$$

$$\frac{\tilde{c}^2}{n^2}\sum_{t=1}^{n}\varphi''(u_t)(y_{t-1}-\tilde{\gamma}'x_{t-1})^2 \Rightarrow -\delta\eta_0^2\int \underline{W}_{\xi_0}(r)^2dr,$$

$$\sum_{t=1}^{n}\varphi'(u_t)(\tilde{\gamma}-\gamma)'\Delta x_{t-1} \Rightarrow \xi_0'\int g(r)dB_\varphi(r),$$

$$\sum_{t=1}^{n}\varphi'(u_t)(\hat{\beta}-\beta)'\breve{x}_t \Rightarrow \int dB_\varphi \underline{X}'\left[\int \underline{X}\underline{X}'dr\right]^{-1}\int \underline{X}dS_0,$$

$$\sum_{t=1}^{n}\varphi''(u_t)[(\tilde{\gamma}-\gamma)'\Delta x_t]^2 \Rightarrow -\delta\xi_0'\int g(r)g(r)'dr\xi_0,$$

$$\sum_{t=1}^{n}\varphi''(u_t)[(\hat{\beta}-\beta)'\breve{x}_t]^2 \Rightarrow -\delta\int dS_0\underline{X}'\left[\int \underline{X}\underline{X}'dr\right]^{-1}\int \underline{X}dS_0.$$

Thus

$$2(L(0,\hat{\gamma})-L(\tilde{c},\tilde{\gamma}))$$

converges weakly to

$$2\eta_0\int \underline{W}_{\xi_0}(r)dB_\varphi(r)+2\delta\eta_0\xi_0'\int g(r)\underline{W}_{\xi_0}(r)dr-\delta\eta_0^2\int \underline{W}_{\xi_0}(r)^2dr$$

$$+2\xi_0'\int g(r)dB_\varphi(r)-\delta\xi_0'\int g(r)g(r)'dr\xi_0$$

$$-2\int dB_\varphi(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_0(r)$$

$$-\delta\int dS_0(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_0(r),$$

and the preceding summation equals

$$\delta\int[\eta_0\underline{W}_{\xi_0}(r)+\xi_0'g(r)]^2dr-2\delta\int[\eta_0\underline{W}_{\xi_0}(r)+\xi_0'g(r)]dS_0(r)$$

$$+\delta\int dS_0(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)dS_0(r). \quad \blacksquare$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

Proof of Theorem 3. The proof follows similar steps to that of Theorem 2. Notice that under the local alternative hypothesis and the given assumptions, the likelihood ratio $L(0,\hat{\gamma}) - L(\tilde{c},\tilde{\gamma})$ has the following asymptotic expansion:

$$\begin{aligned}
&\frac{(\tilde{c}-c)}{n}\sum_{t=1}^{n}\varphi'(u_t)(y_{t-1}-\tilde{\gamma}'x_{t-1}) - \frac{(\tilde{c}-c)}{n}\sum_{t=1}^{n}\varphi''(u_t)(y_{t-1}-\tilde{\gamma}'x_{t-1})(\tilde{\gamma}-\gamma)'\Delta_c x_t \\
&\quad - \frac{(\tilde{c}-c)^2}{2n^2}\sum_{t=1}^{n}\varphi''(u_t)(y_{t-1}-\tilde{\gamma}'x_{t-1})^2 + \sum_{t=1}^{n}\varphi'(u_t)(\tilde{\gamma}-\gamma)'\Delta_c x_{t-1} \\
&\quad + \sum_{t=1}^{n}\varphi'(u_t)\frac{c}{n}y_{t-1}^{s} - \sum_{t=1}^{n}\varphi'(u_t)(\hat{\beta}-\beta)'\breve{x}_t - \frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)[(\tilde{\gamma}-\gamma)'\Delta_c x_t]^2 \\
&\quad - \frac{c}{n}\sum_{t=1}^{n}\varphi''(u_t)y_{t-1}^{s}(\hat{\beta}-\beta)'\breve{x}_t + \frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)[(\hat{\beta}-\beta)'\breve{x}_t]^2 \\
&\quad + \frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left(\frac{c}{n}y_{t-1}^{s}\right)^2 + o_p(1). \qquad \textbf{(A.7)}
\end{aligned}$$

By a calculation of limits of the components in (A.7), we obtain the results of Theorem 3. ■

Proof of Theorem 5. Notice that

$$\begin{aligned}
L(\bar{c},\bar{\gamma}) &= \sum \varphi(\Delta_{\bar{c}}y_t - \bar{\gamma}'\Delta_{\bar{c}}x_t) \\
&= \sum \varphi(u_t - (\bar{c}-c)(y_{t-1}^{s}/n) - (\bar{\gamma}-\gamma)'\Delta_{\bar{c}}x_t);
\end{aligned}$$

under Assumption B, for all c in a compact set, the likelihood ratio $L(0,\hat{\gamma}) - L(\bar{c},\bar{\gamma})$ has the following asymptotic expansion:

$$\begin{aligned}
&\sum_{t=1}^{n}\frac{\varphi'(u_t)}{\sqrt{n}}\left[\frac{cy_{t-1}^{s}}{\sqrt{n}} - \sqrt{n}(\hat{\beta}-\beta)'\breve{x}_t\right] \\
&\quad + \frac{1}{2n}\sum_{t=1}^{n}\varphi''(u_t)\left[\frac{cy_{t-1}^{s}}{\sqrt{n}} - \sqrt{n}(\hat{\beta}-\beta)'\breve{x}_t\right]^2 \\
&\quad - \sum_{t=1}^{n}\varphi'(u_t)\left[\frac{(c-\bar{c})y_{t-1}^{s}}{n} - (\bar{\gamma}-\gamma)'\Delta_{\bar{c}}x_t\right] \\
&\quad - \frac{1}{2}\sum_{t=1}^{n}\varphi''(u_t)\left(\frac{(c-\bar{c})y_{t-1}^{s}}{n} - (\bar{\gamma}-\gamma)'\Delta_{\bar{c}}x_t\right)^2 \\
&\quad + o_p(1).
\end{aligned}$$

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press

It can be verified that

$$\sum_{t=1}^{n} \varphi'(u_t)\,\frac{c}{n}\,y_{t-1}^{s} = c\sum_{t=1}^{n}\frac{y_{t-1}^{s}}{\sqrt{n}}\,\frac{\varphi'(u_t)}{\sqrt{n}} \Rightarrow c\int J_c\,dB_\varphi,$$

$$\sum_{t=1}^{n} \varphi'(u_t)(\hat{\beta}-\beta)'\check{x}_t = [(\hat{\beta}-\beta)'G_n n^{1/2}]\left[\sum_{t=1}^{n} G_n^{-1}\check{x}_t\,\frac{\varphi'(u_t)}{\sqrt{n}}\right]$$

$$\Rightarrow \int dS_c\,\underline{X}'\left[\int \underline{X}\underline{X}'\right]^{-1}\int \underline{X}\,dB_\varphi,$$

$$\frac{c}{n}\sum_{t=1}^{n}\varphi''(u_t)y_{t-1}^{s}(\hat{\beta}-\beta)'\check{x}_t = c\left\{\frac{1}{n}\sum_{t=1}^{n}\varphi''(u_t)\,\frac{y_{t-1}^{s}}{\sqrt{n}}\,\check{x}_t'G_n^{-1}\right\}\{n^{1/2}G_n(\hat{\beta}-\beta)\}$$

$$\Rightarrow -c\delta\int J_c(r)\underline{X}(r)'dr\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)\,dS_c(r),$$

$$\sum_{t=1}^{n}\varphi''(u_t)[(\hat{\beta}-\beta)'\check{x}_t]^2 = [(\hat{\beta}-\beta)'G_n n^{1/2}]\,\frac{1}{n}\sum_{t=1}^{n}\varphi''(u_t)G_n^{-1}\check{x}_t\check{x}_t'G_n^{-1}[n^{1/2}G_n(\hat{\beta}-\beta)]$$

$$\Rightarrow -\delta\int dS_c(r)\underline{X}(r)'\left[\int \underline{X}(r)\underline{X}(r)'dr\right]^{-1}\int \underline{X}(r)\,dS_c(r),$$

$$\sum_{t=1}^{n}\varphi''(u_t)\left(\frac{c}{n}\,y_{t-1}^{s}\right)^2 = c^2\,\frac{1}{n}\sum_{t=1}^{n}\varphi''(u_t)\left(\frac{y_{t-1}^{s}}{\sqrt{n}}\right)^2 \Rightarrow -c^2\delta\int J_c(r)^2dr,$$

$$\sum_{t=1}^{n}\varphi'(u_t)\left(\frac{(\bar{c}-c)y_{t-1}^{s}}{n}+(\bar{\gamma}-\gamma)'\Delta_{\bar{c}}x_t\right)$$

$$\Rightarrow \int\left[(\bar{c}-c)J_c + X_{\bar{c}}\left[\int X_{\bar{c}}X_{\bar{c}}'\right]^{-1}\int X_{\bar{c}}\,d\bar{S}_c\right]dB_\varphi,$$

$$\sum_{t=1}^{n}\varphi''(u_t)\left(\frac{(\bar{c}-c)y_{t-1}^{s}}{n}+(\bar{\gamma}-\gamma)'\Delta_{\bar{c}}x_t\right)^2$$

$$\Rightarrow -\delta\int\left[(\bar{c}-c)J_c + X_{\bar{c}}'\left[\int X_{\bar{c}}X_{\bar{c}}'\right]^{-1}\int X_{\bar{c}}\,d\bar{S}_c\right]^2 dr.$$

Notice that $B_\varphi(r) = -\delta S_0(r)$ and thus

$$-2\int dS_c\,\underline{X}'\left[\int \underline{X}\underline{X}'\right]^{-1}\int \underline{X}\,dB_\varphi - \delta\int dS_c\,\underline{X}'\left[\int \underline{X}\underline{X}'dr\right]^{-1}\int \underline{X}\,dS_c$$

$$= \delta\int dS_c\,\underline{X}'\left[\int \underline{X}\underline{X}'\right]^{-1}\int \underline{X}\,d[2S_0 - S_c]$$

$$= \delta\int dS_c\,\underline{X}'\left[\int \underline{X}\underline{X}'\right]^{-1}\int \underline{X}\,d\underline{S}_c.$$

The result of Theorem 5 can then be obtained. ■

https://doi.org/10.1017/S0266466601176036 Published online by Cambridge University Press